Patent Office
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PO 6974 for a patent by THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH filed on 23 May 1997

I further certify that the annexed specification is not, as yet, open to public inspection.

PRIORITY DOCUMENT

WITNESS my hand this First
day of June 1998

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND SALES

Regulation 3.2

The Council of The Queensland Institute of Medical Research

A U S T R A L I A

Patents Act 1990

PROVISIONAL SPECIFICATION

for the invention entitled:

"A novel gene and uses therefor"

The invention is described in the following statement:

A NOVEL GENE AND USES THEREFOR

The present invention relates generally to a novel human gene and to derivatives and mammalian, animal, avian, insect, nematode, and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined at the end of the description.

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis, conventional pharmaceutical preparations as well as gene and protein replacement therapies.

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. One area of particular interest is in the field of signal transduction.

Knowledge of cellular interaction in the control of cell proliferation is essential in the rational design of specific therapeutic strategies aimed at controlling proliferative disorders. Such proliferative disorders including a range of cancers, inflammatory conditions and atherosclerosis. An important aspect of cellular interaction is in signal transduction *via* receptors to intracellular transducers. One key signal transducer is Ras which couples the

receptors for diverse extracellular signals to different effectors. Ras directly activates the downstream kinase Raf which in turn induces the mitogen activated protein kinase (MAPK) cascade.

The Ras is an example of a guanine nucleotide exchange factor (GEF). A mutation in a GEF such as Ras has been implicated in development of a range of cancers and tumours. There is a need, therefore, to identify new GEFs and to develop therapeutic and diagnostic protocols based on modulating function of the GEF singalling pathways.

Accordingly, one aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative of said gene regulator.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

(i) a nucleotide sequence set forth in SEQ ID NO:1;

(ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:2;

(iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and

(iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative

stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

The present invention extends to nucleic acid molecules with percentage similarities of approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

The nucleic acid molecule of the present invention is hereinafter referred to as constituting the "*mcg7*" gene. The protein encoded by *mcg7* is referred to herein as "MCG7" and is involved in signal transduction.

The present invention extends to the naturally occurring genomic *mcg7* nucleotide sequence or corresponding cDNA sequence or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG7 or the corresponding genetic sequence. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG7 or single or multiple nucleotide substitutions, deletions and/or additions to *mcg7*. Derivatives also includes modifications to nucleotide bases or amino acid residues to, for example, alter glycosylation sites or amino acid side chains. "Additions" to the amino acid or nucleotide sequences include fusions with

other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG7" or "*mcg7*" includes references to all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG7.

The *mcg7* of the present invention is particularly exemplified herein from humans and in particular from human chromosome 11q13.

The present invention also extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), birds (eg. chickens, ducks, geese, parrot), insects, nematodes, eukaryotic microorganisms and captive wild animals (eg. deer, foxes, kangaroos). Reference herein to *mcg7* or MCG7 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an *mcg7* polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg7* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg7* gene portion in an appropriate cell.

In addition, the *mcg7* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG7 is a GEF involved in signal transduction. Mutations in *mcg7* or MCG7 may result in defective control of cell proliferation leading to the development of or a propensity to develop various types of cancer.

A deletion or aberration in the *mcg7* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg7,* said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

The nucleotide substitutions, additions or deletions may be detected by any convenient means including nucleotide sequencing, restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), oligonucleotide hybridization and single stranded

conformation polymorphism analysis (SSCP) amongst many others. An aberration includes modification to existing nucleotides such as to modify glycosylation signals amongst other effects.

In an alternative method, aberrations in the *mcg7* gene are detected by screening for mutations in MCG7.

A mutation in MCG7 may be a single or multiple amino acid substitution, addition and/or deletion. The mutation in *mcg7* may also result in either no translation product being produced or a product in truncated form. A mutation may also be an altered glycosylation pattern or the introduction of side chain modifications to amino acid residues.

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

A particularly convenient means of detecting a mutation in MCG7 is by use of antibodies.

Accordingly another aspect of the present invention is directed to antibodies to MCG7 and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG7 or may be specifically raised to MCG7 or derivatives thereof. In the case of the latter, MCG7 or its derivatives may first need to be associated with a carrier molecule. The antibodies to MCG7 of the present invention are particularly useful as diagnostic agents.

For example, antibodies to MCG7 and its derivatives can be used to screen for wild-type MCG7 or for mutated MCG7 molecules. The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG7 levels or the presence of wild-type MCG7 may be important for

diagnosis of certain cancers or a predisposition for development of cancers or for monitoring certain therapeutic protocols.

As stated above antibodies to MCG7 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

For example, specific antibodies can be used to screen for wild-type MCG7 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for example, as a means for screening for levels of MCG7 in a cell extract or other biological fluid or purifying MCG7 made by recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG7 or to a specific mutant phenotype or to a deleted or otherwise altered region.

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG7 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal or bird with an effective amount of MCG7 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

The presence of MCG7 may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4, 424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing

with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG7 including cell extract or, tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the MCG7 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes) and under suitable conditions (e.g. 25°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are

particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

As stated above, the present invention extends to genetic constructs capable of encoding MCG7 or functional derivatives thereof. Such genetic constructs are also contemplated to be useful in modulating expression of specific genes in which *mcg7* is involved in tissue-specific or temporal regulation.

Accordingly, another aspect of the present invention is directed to a genetic construct comprising a nucleotide sequence encoding a peptide, polypeptide or protein and *mcg7* or a functional derivative or homologue thereof capable of modulating the expression of said nucleotide sequence.

The present invention is further described with reference to the following non-limiting figures and Examples.

In the Figures:

Figure 1 is a representation showing similarity of MCG7 with GEFs of organisms.

Figure 2(a) is a representation of the nucleotide sequence and corresponding amino acid sequence of mcg7. An exon is shown in the nucleotide sequence in lower case (nucleotides 183-298).

Figure 2(b) is a representation of the nucleotide sequence and corresponding amino acid sequence of mcg7 but without the exon shown in Fig. 2(a). The cDNA molecules of Fig. 2(a) and Fig.2(b) differ by the inclusion and exclusion of the exon shown in Figure 2(a) in lower case.

Figure 3 is a representation showing a comparison between MCG7 and a homologue from *Caenorhabditis elegans* using BEST FIT algorithm. This top codon is also present in a mouse EST and sequence alignment between human and mouse ESTs suggests this region represents the 5' UTR. Furhtermore, protein homology with the *C. elegans* protein (shown below) suggests the underlined ATG codon to represent the true initiation codon.

In the figure, the following sequences are colour coded:

orange	1	nematode	DVDEEDEVEDIEF
orange	3	human	DVDGDGHISQEEF
		nematode	DHDRDGFISQEEF

orange	4	human	DQNQDGCISREEM
		nematode	DVDMDGQISKDEL
pink	2	human	HFVHVAEKVVQLQNFNTLMAVVGGLSHSSISRLKETH
		nematode	KFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTY
yellow	5		
human			HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVE
nematode			HNFHETTFLTPTTCNHCNKLLWGILRQGFKCKDCGLAVHSCCKSNAVAE

Figure 4 is a representation of an alignment of human and murine *mcg*7 nucleotide sequences.

Figure 5 is a representation of further 5' nucleotide and corresponding amino acid sequence for *mcg*7.

Figure 6 is a graphical representation of GDP release assay. □ Experiment #1 (mean of duplicates). ◇ Experiment #2 (mean of duplicates). Exchange reaction contained 36pMols of GSTmcg7 (N-terminally truncated) and 1.6-12.8 pMols of recombinant N-Ras.GDP. Reaction time 6 mins.

Estimated reaction constants:

K_m = 2.1μM, V_{max} = 37pMol/6min/36pMol [Expt#1]

K_m = 1.5μM, V_{max} = 30.3pMol/6 min/36pMol [Expt#2]

EXAMPLE 1

A human gene (designated *mcg7*) was identified and isolated from chromosome 11q13 which encodes a protein that bears striking homology with guanine nucleotide exchange factors (GEFs) from a wide variety of organisms (Fig. 1).

EXAMPLE 2

The composite *mcg7* cDNA sequence is at least 2.4kb in length and Figures 2(a) and 2(b) show a predicted amino acid sequence of 609 amino acids. Alternative start sites may yield a protein of 714 amino acids (Fig.5).

EXAMPLE 3

A *mcg7* homologue from *C. elegans* has been identified, the product of which is highly conserved with that of MCG7 (Fig. 3). There are several salient features of the protein which have been highlighted in Fig. 3 - namely: a guanine nucleotide binding region (pink), a diacylglycerol binding region (yellow), and "EF-hand"-calcium binding regions (orange). In addition, there are several potential cAMP, protein kinase C, and casin kinase II phosphorylation sites, as well as a number of potential sites for glycosylation (not indicated).

EXAMPLE 4

A number of partial human and murine EST clones exist for *mcg7*.

EXAMPLE 5

The best characterised GEFs are the family of *ras* oncoproteins, which play a pivotal role in signal transduction and when mutated are responsible for tumour development. A variety of therapeutic regimes for cancer treatment have been designed to specifically interfere with the

ras signalling pathways. There is potential, therefore that the product of *mcg7* could also be a target for such clinical strategies.

EXAMPLE 6

Intitiation codons for mcg7.

The nucleotide sequence for mcg7 cDNA was extended 5' with genomic DNA sequence from Genbank accession number AC000134 (positions 1-321) and analysed for additional coding sequence 5' to the putative initiation codon (nt 681-683). An additional in-frame ATG occurs at position nt 495-497 when the alternatively splice exon (position nt 504-609) is present. This closely matches the Kozak consensus. When this exon is absent, then the ATG is not in-frame and other possible initiation codons are absent resulting translation shown in lower case lettering. Further evidence that the initiation codon at position nt 681-683 is the true intitation site is given below in Figure 4.

Alignment of human and murine mcg7 cDNA sequences is shown in Figure 4. The murin sequence represents a composite of 2 cDNA sequences from the expressed sequence tag database (accession numbers W71787 and AA237373). The putative initiation codon is at position nt 360-362. Both murine ESTs appear to have an upstream in-frame stop codon at position nt 326-328, downstream of the differentially spliced exon. Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

The data are shown in Figures 4 and 5 and strongly suggest that the ATG codon at position nt 360-362 encodes the N-terminus of MCG7.

EXAMPLE 7

Figure 6 shows data from experiments indicating that a truncated version of mcg7 when expressed as a GST fursion protein can function as a Ras guanine nucleotide exchange factor. In brief, Ras (unprocessed) is loaded with ^{3}H-GDP then incubated in the presence of excess

cold GTP ± GSTmcg7. Full details of our assay can be found in Porfiri et al. J. Biol. Chem. 269, 22672-22677 (1994).

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: The Council of The Queensland Institute for Medical Research

(ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: AUSTRALIAN PROVISIONAL
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: HUGHES, DR E JOHN L
(C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: +61 3 9254 2777
(B) TELEFAX: +61 3 9254 2770
(C) TELEX: AA 31787

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..2188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC TCC CCA AAA TAT TCC          47
   Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser
    1               5                  10                  15

CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG CTG TCC          95
His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser
                20                  25                  30

CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG CCT GCA GGA         143
Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly
            35                  40                  45

AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA ACG CAG CGC CTG         191
Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu
        50                  55                  60

TGT GGC CGC GGG ACT CAA GGC TGG CCT GGC TCA AGT GAA CAG CAC GTC         239
Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val
    65                  70                  75

CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT TCT GGG GTG GAC GAG CTG         287
Gln Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu
80                  85                  90                  95

GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC         335
Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly
                100                 105                 110

CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC CTG GAC         383
Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp
            115                 120                 125

AAG GGC TGC ACG GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC         431
Lys Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe
        130                 135                 140

GAT GAC TCC GGG AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC         479
Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu
    145                 150                 155

ATG ATG CAC CCC TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG         527
Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu
160                 165                 170                 175

CTC CAC ATC TAC CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG         575
Leu His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln
                180                 185                 190

GTG AAA ACG TGC CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG         623
Val Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala
            195                 200                 205

GAG TTT GAC TTG AAC CCG GAG TTG GCT GAG CAG ATC AAG GAG CTG AAG         671
Glu Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys
        210                 215                 220
```

```
GCT CTG CTA GAC CAA GAA GGG AAC CGA CGG CAC AGC AGC CTA ATC GAC        719
Ala Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp
    225                 230                 235

ATA GAC AGC GTC CCT ACC TAC AAG TGG AAG CGG CAG GTG ACT CAG CGG        767
Ile Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg
240                 245                 250                 255

AAC CCT GTG GGA CAG AAA AAG CGC AAG ATG TCC CTG TTG TTT GAC CAC        815
Asn Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His
                260                 265                 270

CTG GAG CCC ATG GAG CTG GCG GAG CAT CTC ACC TAC TTG GAG TAT CGC        863
Leu Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg
            275                 280                 285

TCC TTC TGC AAG ATC CTG TTT CAG GAC TAT CAC AGT TTC GTG ACT CAT        911
Ser Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His
        290                 295                 300

GGC TGC ACT GTG GAC AAC CCC GTC CTG GAG CGG TTC ATC TCC CTC TTC        959
Gly Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe
    305                 310                 315

AAC AGC GTC TCA CAG TGG GTG CAG CTC ATG ATC CTC AGC AAA CCC ACA       1007
Asn Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr
320                 325                 330                 335

GCC CCG CAG CGG GCC CTG GTC ATC ACA CAC TTT GTC CAC GTG GCG GAG       1055
Ala Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu
                340                 345                 350

AAG CTG CTA CAG CTG CAG AAC TTC AAC ACG CTG ATG GCA GTG GTC GGG       1103
Lys Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly
            355                 360                 365

GGC CTG AGC CAC AGC TCC ATC TCC CGC CTC AAG GAG ACC CAC AGC CAC       1151
Gly Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His
        370                 375                 380

GTT AGC CCT GAG ACC ATC AAG CTC TGG GAG GGT CTC ACG GAA CTA GTG       1199
Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val
    385                 390                 395

ACG GCG ACA GGC AAC TAT GGC AAC TAC CGG CGT CGG CTG GCA GCC TGT       1247
Thr Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys
400                 405                 410                 415

GTG GGC TTC CGC TTC CCG ATC CTG GGT GTG CAC CTC AAG GAC CTG GTG       1295
Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val
                420                 425                 430

GCC CTG CAG CTG GCA CTG CCT GAC TGG CTG GAC CCA GCC CGG ACC CGG       1343
Ala Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg
            435                 440                 445

CTC AAC GGG GCC AAG ATG AAG CAG CTC TTT AGC ATC CTG GAG GAG CTG       1391
Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu
        450                 455                 460

GCC ATG GTG ACC AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG       1439
Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu
    465                 470                 475
```

```
CTG AGC CTG CTC ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG      1487
Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu
480                 485                 490                 495

CTG TAC CAG CTG TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA      1535
Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro
                500                 505                 510

ACC AGC CCC ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG      1583
Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu
            515                 520                 525

GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG      1631
Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val
        530                 535                 540

GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC      1679
Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val
    545                 550                 555

GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG      1727
Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly
560                 565                 570                 575

AAC TTC CCT TAC CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT      1775
Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp
                580                 585                 590

GGC TGC ATC AGC AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC      1823
Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser
            595                 600                 605

TCT GTG TTG GGG GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC      1871
Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser
        610                 615                 620

AAC TCC TTG CGC CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG      1919
Asn Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu
    625                 630                 635

GGC ATC TAC AAG CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC      1967
Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys
640                 645                 650                 655

CAC AAG CAG TGC AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC      2015
His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala
                660                 665                 670

CAG AGT GTG AGC CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC      2063
Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His
            675                 680                 685

AGC CAC CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG      2111
Ser His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg
        690                 695                 700

CGA GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG      2159
Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val
    705                 710                 715

GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG TGGTTGGATC        2208
Glu Asp Gly Val Phe Asp Ile His Leu
720                 725
```

AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA GCAGGGAGCC TGGGGGTGTC 2268

GGGGCAGGAG GCTGGGGATG GGGGTGGGAT ATGAGGGTGG CATGCAGCTG AGGGCAGGGC 2328

CAGGGCTGGT GTCCCTAAGG TTGTACAGAC TCTTGTGAAT ATTTGTATTT TCCAGATGGA 2388

ATAAAAAGGC CCGTGTAATT AACCTTC 2415

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 728 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser His
 1               5                  10                  15

Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser Pro
            20                  25                  30

Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly Arg
        35                  40                  45

Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu Cys
    50                  55                  60

Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val Gln
65                  70                  75                  80

Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu Gly
                85                  90                  95

Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro
            100                 105                 110

Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys
        115                 120                 125

Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp
    130                 135                 140

Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met
145                 150                 155                 160

Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu
                165                 170                 175

His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val
            180                 185                 190

Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu
        195                 200                 205

Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala
    210                 215                 220

Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile
225                 230                 235                 240
```

Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn
245 250 255

Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu
260 265 270

Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser
275 280 285

Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His Gly
290 295 300

Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe Asn
305 310 315 320

Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr Ala
325 330 335

Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu Lys
340 345 350

Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly Gly
355 360 365

Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His Val
370 375 380

Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr
385 390 395 400

Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val
405 410 415

Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala
420 425 430

Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu
435 440 445

Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala
450 455 460

Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu
465 470 475 480

Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu
485 490 495

Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr
500 505 510

Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu
515 520 525

Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu
530 535 540

His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp
545 550 555 560

Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn
565 570 575

Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly
580 585 590

```
Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser
        595                 600                 605

Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn
    610                 615                 620

Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly
625                 630                 635                 640

Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His
                645                 650                 655

Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln
            660                 665                 670

Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser
        675                 680                 685

His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg
    690                 695                 700

Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu
705                 710                 715                 720

Asp Gly Val Phe Asp Ile His Leu
                725
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 300 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 170..300

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG      60

CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC     120

TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAG CCC CAT         175
                                                          Pro His
                                                            1

GGG AAC GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC       223
Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser
          5                  10                  15

CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC       271
Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp
     20                  25                  30

CTG GAC AAG GGC TGC ACG GTG GAG GAG CT                               300
Leu Asp Lys Gly Cys Thr Val Glu Glu Leu
 35                  40
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Pro His Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu
  1               5                  10                  15

Arg Ser Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr
             20                  25                  30

Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu
         35                  40
```

DATED this 23rd day of May 1997

The Council of The Queensland Institute for Medical Research

By DAVIES COLLISON CAVE

Patent Attorneys for the Applicants

FIGURE 2

```
CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC TCC CCA AAA TAT TCC          47
   Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser
     1               5                  10                  15

CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG CTG TCC          95
His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser
                 20                  25                  30

CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG CCT GCA GGA         143
Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly
            35                  40                  45

AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA ACG CAG CGC CTG         191
Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu
        50                  55                  60

TGT GGC CGC GGG ACT CAA GGC TGG CCT GGC TCA AGT GAA CAG CAC GTC         239
Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val
    65                  70                  75

CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT TCT GGG GTG GAC GAG CTG         287
Gln Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu
80                  85                  90                  95

GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC         335
Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly
                100                 105                 110

CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC CTG GAC         383
Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp
            115                 120                 125

AAG GGC TGC ACG GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC         431
Lys Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe
        130                 135                 140

GAT GAC TCC GGG AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC         479
Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu
    145                 150                 155

ATG ATG CAC CCC TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG         527
Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu
160                 165                 170                 175

CTC CAC ATC TAC CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG         575
Leu His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln
                180                 185                 190

GTG AAA ACG TGC CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG         623
Val Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala
            195                 200                 205

GAG TTT GAC TTG AAC CCG GAG TTG GCT GAG CAG ATC AAG GAG CTG AAG         671
```

Figure 2 (continued)

```
Glu Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys
        210                 215                 220

GCT CTG CTA GAC CAA GAA GGG AAC CGA CGG CAC AGC AGC CTA ATC GAC         719
Ala Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp
    225                 230                 235
```

```
ATA GAC AGC GTC CCT ACC TAC AAG TGG AAG CGG CAG GTG ACT CAG CGG        767
Ile Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg
240                 245                 250                 255

AAC CCT GTG GGA CAG AAA AAG CGC AAG ATG TCC CTG TTG TTT GAC CAC        815
Asn Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His
                260                 265                 270

CTG GAG CCC ATG GAG CTG GCG GAG CAT CTC ACC TAC TTG GAG TAT CGC        863
Leu Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg
            275                 280                 285

TCC TTC TGC AAG ATC CTG TTT CAG GAC TAT CAC AGT TTC GTG ACT CAT        911
Ser Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His
        290                 295                 300

GGC TGC ACT GTG GAC AAC CCC GTC CTG GAG CGG TTC ATC TCC CTC TTC        959
Gly Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe
    305                 310                 315

AAC AGC GTC TCA CAG TGG GTG CAG CTC ATG ATC CTC AGC AAA CCC ACA       1007
Asn Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr
320                 325                 330                 335

GCC CCG CAG CGG GCC CTG GTC ATC ACA CAC TTT GTC CAC GTG GCG GAG       1055
Ala Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu
                340                 345                 350

AAG CTG CTA CAG CTG CAG AAC TTC AAC ACG CTG ATG GCA GTG GTC GGG       1103
Lys Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly
            355                 360                 365

GGC CTG AGC CAC AGC TCC ATC TCC CGC CTC AAG GAG ACC CAC AGC CAC       1151
Gly Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His
        370                 375                 380

GTT AGC CCT GAG ACC ATC AAG CTC TGG GAG GGT CTC ACG GAA CTA GTG       1199
Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val
    385                 390                 395

ACG GCG ACA GGC AAC TAT GGC AAC TAC CGG CGT CGG CTG GCA GCC TGT       1247
Thr Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys
400                 405                 410                 415

GTG GGC TTC CGC TTC CCG ATC CTG GGT GTG CAC CTC AAG GAC CTG GTG       1295
Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val
                420                 425                 430

GCC CTG CAG CTG GCA CTG CCT GAC TGG CTG GAC CCA GCC CGG ACC CGG       1343
Ala Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg
```

Figure 2 (continued)

```
            435                 440                 445

CTC AAC GGG GCC AAG ATG AAG CAG CTC TTT AGC ATC CTG GAG GAG CTG       1391
Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu
        450                 455                 460

GCC ATG GTG ACC AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG       1439
Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu
    465                 470                 475

CTG AGC CTG CTC ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG       1487
Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu
480                 485                 490                 495

CTG TAC CAG CTG TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA       1535
Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro
                500                 505                 510
```

```
ACC AGC CCC ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG      1583
Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu
            515                 520                 525

GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG      1631
Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val
        530                 535                 540

GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC      1679
Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val
    545                 550                 555

GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG      1727
Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly
560                 565                 570                 575

AAC TTC CCT TAC CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT      1775
Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp
                580                 585                 590

GGC TGC ATC AGC AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC      1823
Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser
            595                 600                 605

TCT GTG TTG GGG GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC      1871
Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser
        610                 615                 620

AAC TCC TTG CGC CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG      1919
Asn Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu
    625                 630                 635

GGC ATC TAC AAG CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC      1967
Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys
640                 645                 650                 655

CAC AAG CAG TGC AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC      2015
His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala
```

Figure 2 (continued)

```
                660                 665                 670

CAG AGT GTG AGC CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC      2063
Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His
            675                 680                 685

AGC CAC CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG      2111
Ser His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg
        690                 695                 700

CGA GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG      2159
Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val
    705                 710                 715

GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG TGGTTGGATC         2208
Glu Asp Gly Val Phe Asp Ile His Leu
720                 725

AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA GCAGGGAGCC TGGGGGTGTC    2268

GGGGCAGGAG GCTGGGGATG GGGGTGGGAT ATGAGGGTGG CATGCAGCTG AGGGCAGGGC    2328

CAGGGCTGGT GTCCCTAAGG TTGTACAGAC TCTTGTGAAT ATTTGTATTT TCCAGATGGA    2388

ATAAAAAGGC CCGTGTAATT AACCTTCA                                       2416
```

Figure 1

Sequences producing High-scoring Segment Pairs:		High Score	Smallest Sum Probability P(N)	N
gnl|PID|e236178	(Z70752) F25B3.3 [Caenorhabditis ele...	307	3.0e-124	8
gi|1293099	(U53884) aimless RasGEF [Dictyosteli...	202	7.8e-22	5
gi|1655941	(U67326) Ras-GRF2 [Mus musculus]	152	3.6e-16	4
pir||S30356	CDC25 protein homolog - yeast (Candi...	150	2.2e-15	3
sp|P43069|CC25_CANAL	CELL DIVISION CONTROL PROTEIN 25	150	2.2e-15	3
sp|P28818|GNRP_RAT	GUANINE NUCLEOTIDE RELEASING PROTEIN...	166	2.6e-15	3
prf||1814463A	guanine nucleotide-releasing factor ...	166	2.6e-15	3
pir||B46199	nucleotide-exchange-factor homolog c...	167	1.1e-14	1
gnl|PID|e238680	(X97560) hypothetical protein L1309 ...	158	3.0e-14	3
pir||S22693	CDC25 protein homolog - mouse /gi|50...	167	3.7e-14	2
sp|P14771|SC25_YEAST	SCD25 PROTEIN /gi|457494 (M26647) SD...	158	4.6e-14	3
sp|P26674|STE6_SCHPO	STE6 PROTEIN /pir||S28098 ste6 prote...	160	5.2e-14	2
pir||S28407	CDC25 protein homolog - mouse	167	1.2e-13	3
sp|P27671|GNRP_MOUSE	GUANINE NUCLEOTIDE RELEASING PROTEIN...	167	1.2e-13	3
gi|386047	(S62035) Ras-specific guanine nucleo...	153	2.0e-13	2
sp|Q02342|CC25_SACKL	CELL DIVISION CONTROL PROTEIN 25 /pi...	142	4.5e-13	2
pir||S14177	SCD25 protein - yeast (Saccharomyces...	152	5.7e-13	3
gi|433720	(L26584) CDC25 [Homo sapiens]	153	6.0e-13	3
gnl|PID|e241744	(Z68880) T14G10.2 [Caenorhabditis el...	157	7.2e-13	1
gi|3484	(X03579) CDC25 protein (aa 1-1588) [...	136	3.4e-12	3
sp|P04821|CC25_YEAST	CELL DIVISION CONTROL PROTEIN 25 /pi...	136	3.4e-12	3
gi|915328	(U24070) Munc13-1 [Rattus norvegicus]	151	5.5e-12	1
pir||A46199	nucleotide-exchange-factor homolog c...	149	5.6e-12	1
pdb|1PTR|	Molecule: Protein Kinase C Delta Ty...	136	1.5e-11	1
gi|915330	(U24071) Munc13-2 [Rattus norvegicus]	150	1.6e-11	2
gi|474982	(D21239) 'C3G protein' [Homo sapiens...	131	3.3e-11	3
gi|1763306	(U75361) Munc13-3 [Rattus norvegicus]	153	6.4e-11	2
gi|806957	guanine-nucleotide exchange factor C...	128	7.8e-11	3
sp|Q03385|GNDS_MOUSE	GUANINE NUCLEOTIDE DISSOCIATION STIM...	133	1.0e-10	2
pir||BVBYL1	LTE1 protein - yeast (Saccharomyces ...	139	1.9e-10	1
gi|452242	(D21354) a putative guanine nucleoti...	139	2.7e-10	1
sp|P07866|LTE1_YEAST	LOW TEMPERATURE ESSENTIAL PROTEIN /p...	139	2.7e-10	1
gi|509050	(Z22521) protein kinase C delta [Hom...	137	4.0e-10	1
gi|520587	(D10495) protein kinase C delta-type...	137	4.6e-10	1
sp|P05130|KPC1_DROME	PROTEIN KINASE C, BRAIN ISOZYME (PKC...	137	4.7e-10	1
pir||S35704	protein kinase C (EC 2.7.1.-) delta ...	137	4.7e-10	1
sp|Q05655|KPCD_HUMAN	PROTEIN KINASE C, DELTA TYPE (NPKC-D...	137	4.7e-10	1
pir||S40279	protein kinase C mu - human /pir||A5...	137	4.9e-10	1
sp|P09215|KPCD_RAT	PROTEIN KINASE C, DELTA TYPE (NPKC-D...	135	9.0e-10	1
gi|520878	(Z34524) serine/threonine protein ki...	133	1.8e-09	1
gi|1519719	(U68142) RalGDS-like [Homo sapiens]	115	3.8e-09	3

FIGURE 2

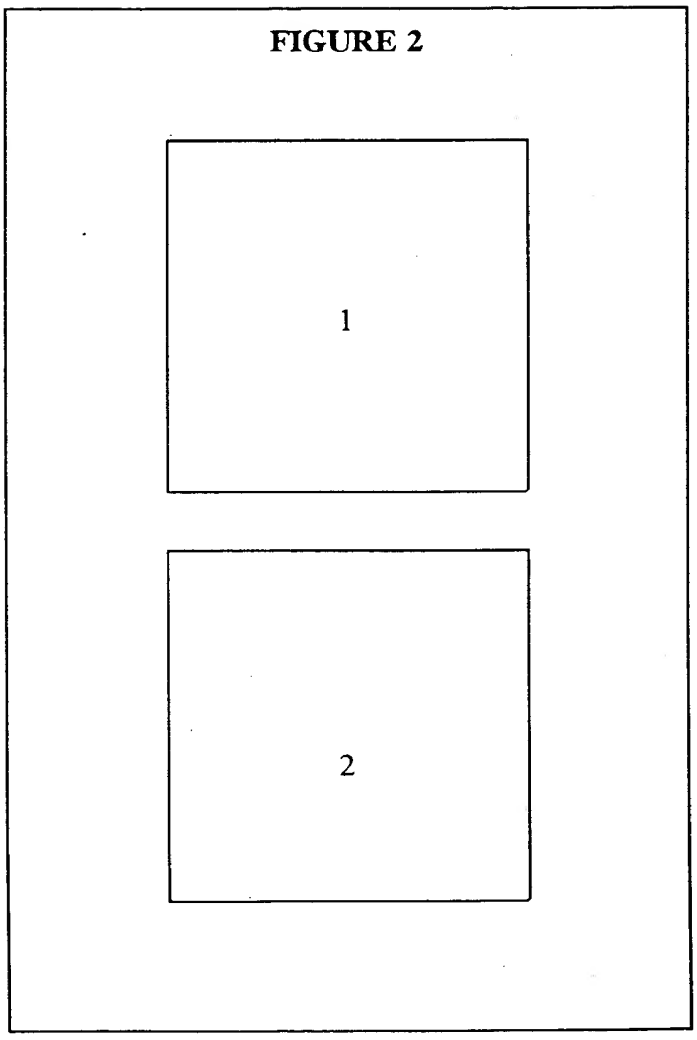


Figure 2a

MCG7 - Cloning of a novel human gene that encodes a guanine exchange factor

```
CGATTTCATTCCTCGCTCCCCACAGGTCCCTCTCCCCAAAATATTCCCATCTTGTCCTAG 60
   I  S  F  L  A  P  H  R  S  L  S  P  K  Y  S  H  L  V  L   19
CCCATCCCCCAGACTATCTCAAGGACCAGCTGTCCCCACGCCCCCGACCTCCACTAGGCC 120
A  H  P  P  D  Y  L  K  D  Q  L  S  P  R  P  R  P  P  L  G   39
TGTGCCACCCGCTGCCTGCAGGAAGACGCCCGGTCCCGGGCCGGGTTAGCCCCATGGGAA 180
L  C  H  P  L  P  A  G  R  R  P  V  P  G  R  V  S  P  M  G   59
CGcagcgcctgtgtggccgcgggactcaaggctggcctggctcaagtgaacagcacgtcc 240
T  Q  R  L  C  G  R  G  T  Q  G  W  P  G  S  S  E  Q  H  V   79
aggaggcgacctcgtccgcgggtttgcattctggggtggacgagctggGGGTTCGGTCCG 300
Q  E  A  T  S  S  A  G  L  H  S  G  V  D  E  L  G  V  R  S   99
AGCCCGGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCAGCCCACCCCGCGCCGGCGGCCA 360
E  P  G  G  R  L  P  E  R  S  L  G  P  A  H  P  A  P  A  A  119
TGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGGAGGAGCTGCTCCGCGGGTGCA 420
M  A  G  T  L  D  L  D  K  G  C  T  V  E  E  L  L  R  G  C  139
TCGAAGCCTTCGATGACTCCGGGAAGGTGCGGGACCCGCAGCTGGTGCGCATGTTCCTCA 480
I  E  A  F  D  D  S  G  K  V  R  D  P  Q  L  V  R  M  F  L  159
TGATGCACCCCTGGTACATCCCCTCCTCTCAGCTGGCGGCCAAGCTGCTCCACATCTACC 540
M  M  H  P  W  Y  I  P  S  S  Q  L  A  A  K  L  L  H  I  Y  179
AACAATCCCGGAAGGACAACTCCAATTCCCTGCAGGTGAAAACGTGCCACCTGGTCAGGT 600
Q  Q  S  R  K  D  N  S  N  S  L  Q  V  K  T  C  H  L  V  R  199
ACTGGATCTCCGCCTTCCCAGCGGAGTTTGACTTGAACCCGGAGTTGGCTGAGCAGATCA 660
Y  W  I  S  A  F  P  A  E  F  D  L  N  P  E  L  A  E  Q  I  219
AGGAGCTGAAGGCTCTGCTAGACCAAGAAGGGAACCGACGGCACAGCAGCCTAATCGACA 720
K  E  L  K  A  L  L  D  Q  E  G  N  R  R  H  S  S  L  I  D  239
TAGACAGCGTCCCTACCTACAAGTGGAAGCGGCAGGTGACTCAGCGGAACCCTGTGGGAC 780
I  D  S  V  P  T  Y  K  W  K  R  Q  V  T  Q  R  N  P  V  G  259
AGAAAAAGCGCAAGATGTCCCTGTTGTTTGACCACCTGGAGCCCATGGAGCTGGCGGAGC 840
Q  K  K  R  K  M  S  L  L  F  D  H  L  E  P  M  E  L  A  E  279
ATCTCACCTACTTGGAGTATCGCTCCTTCTGCAAGATCCTGTTTCAGGACTATCACAGTT 900
H  L  T  Y  L  E  Y  R  S  F  C  K  I  L  F  Q  D  Y  H  S  299
TCGTGACTCATGGCTGCACTGTGGACAACCCCGTCCTGGAGCGGTTCATCTCCCTCTTCA 960
F  V  T  H  G  C  T  V  D  N  P  V  L  E  R  F  I  S  L  F  319
ACAGCGTCTCACAGTGGGTGCAGCTCATGATCCTCAGCAAACCCACAGCCCCGCAGCGGG 1020
N  S  V  S  Q  W  V  Q  L  M  I  L  S  K  P  T  A  P  Q  R  339
CCCTGGTCATCACACACTTTGTCCACGTGGCGGAGAAGCTGCTACAGCTGCAGAACTTCA 1080
A  L  V  I  T  H  F  V  H  V  A  E  K  L  L  Q  L  Q  N  F  359
ACACGCTGATGGCAGTGGTCGGGGGCCTGAGCCACAGCTCCATCTCCCGCCTCAAGGAGA 1140
N  T  L  M  A  V  V  G  G  L  S  H  S  S  I  S  R  L  K  E  379
CCCACAGCCACGTTAGCCCTGAGACCATCAAGCTCTGGGAGGGTCTCACGGAACTAGTGA 1200
T  H  S  H  V  S  P  E  T  I  K  L  W  E  G  L  T  E  L  V  399
CGGCGACAGGCAACTATGGCAACTACCGGCGTCGGCTGGCAGCCTGTGTGGGCTTCCGCT 1260
T  A  T  G  N  Y  G  N  Y  R  R  R  L  A  A  C  V  G  F  R  419
TCCCGATCCTGGGTGTGCACCTCAAGGACCTGGTGGCCCTGCAGCTGGCACTGCCTGACT 1320
F  P  I  L  G  V  H  L  K  D  L  V  A  L  Q  L  A  L  P  D  439
GGCTGGACCCAGCCCGGACCCGGCTCAACGGGGCCAAGATGAAGCAGCTCTTTAGCATCC 1380
W  L  D  P  A  R  T  R  L  N  G  A  K  M  K  Q  L  F  S  I  459
TGGAGGAGCTGGCCATGGTGACCAGCCTGCGGCCACCAGTACAGGCCAACCCCGACCTGC 1440
L  E  E  L  A  M  V  T  S  L  R  P  P  V  Q  A  N  P  D  L  479
TGAGCCTGCTCACGGTGTCTCTGGATCAGTATCAGACGGAGGATGAGCTGTACCAGCTGT 1500
L  S  L  L  T  V  S  L  D  Q  Y  Q  T  E  D  E  L  Y  Q  L  499
CCCTGCAGCGGGAGCCGCGCTCCAAGTCCTCGCCAACCAGCCCCACGAGTTGCACCCCAC 1560
S  L  Q  R  E  P  R  S  K  S  S  P  T  S  P  T  S  C  T  P  519
CACCCCGGCCCCCGGTACTGGAGGAGTGGACCTCGGCTGCCAAACCCAAGCTGGATCAGG 1620
P  P  R  P  P  V  L  E  E  W  T  S  A  A  K  P  K  L  D  Q  539
CCCTCGTGGTGGAGCACATCGAGAAGATGGTGGAGTCTGTGTTCCGGAACTTTGACGTCG 1680
```

Figure 2a (cont...)

```
A  L  V  V  E  H  I  E  K  M  V  E  S  V  F  R  N  F  D  V   559
ATGGGGATGGCCACATCTCACAGGAAGAATTCCAGATCATCCGTGGGAACTTCCCTTACC 1740
D  G  D  G  H  I  S  Q  E  E  F  Q  I  I  R  G  N  F  P  Y   579
TCAGCGCCTTTGGGGACCTCGACCAGAACCAGGATGGCTGCATCAGCAGGGAGGAGATGG 1800
L  S  A  F  G  D  L  D  Q  N  Q  D  G  C  I  S  R  E  E  M   599
TTTCCTATTTCCTGCGCTCCAGCTCTGTGTTGGGGGGGCGCATGGGCTTCGTACACAACT 1860
V  S  Y  F  L  R  S  S  S  V  L  G  G  R  M  G  F  V  H  N   619
TCCAGGAGAGCAACTCCTTGCGCCCCGTCGCCTGCCGCCACTGCAAAGCCCTGATCCTGG 1920
F  Q  E  S  N  S  L  R  P  V  A  C  R  H  C  K  A  L  I  L   639
GCATCTACAAGCAGGGCCTCAAATGCCGAGCCTGTGGAGTGAACTGCCACAAGCAGTGCA 1980
G  I  Y  K  Q  G  L  K  C  R  A  C  G  V  N  C  H  K  Q  C   659
AGGATCGCCTGTCAGTTGAGTGTCGGCGCAGGGCCCAGAGTGTGAGCCTGGAGGGGTCTG 2040
K  D  R  L  S  V  E  C  R  R  R  A  Q  S  V  S  L  E  G  S   679
CACCCTCACCCTCACCCATGCACAGCCACCATCACCGCGCCTTCAGCTTCTCTCTGCCCC 2100
A  P  S  P  S  P  M  H  S  H  H  H  R  A  F  S  F  S  L  P   699
GCCCTGGCAGGCGAGGCTCCAGGCCTCCAGAGATCCGTGAGGAGGAGGTACAGACGGTGG 2160
R  P  G  R  R  G  S  R  P  P  E  I  R  E  E  E  V  Q  T  V   719
AGGATGGGGTGTTTGACATCCACTTGTAATAGATGCTGTGGTTGGATCAAGGACTCATTC 2220
E  D  G  V  F  D  I  H  L  *                                  728
CTGCCTTGGAGAAAATACTTCAACCAGAGCAGGGAGCCTGGGGGTGTCGGGGCAGGAGGC 2280
TGGGGATGGGGGTGGGATATGAGGGTGGCATGCAGCTGAGGGCAGGGCCAGGGCTGGTGT 2340
CCCTAAGGTTGTACAGACTCTTGTGAATATTTGTATTTTCCAGATGGAATAAAAAGGCCC 2400

GTGTAATTAACCTTC(A)n
```

Figure 2b

```
CGATTTCATTCCTCGCTCCCCACAGGTCCCTCTCCCCAAAATATTCCCATCTTGTCCTAG 60
CCCATCCCCCAGACTATCTCAAGGACCAGCTGTCCCCACGCCCCCGACCTCCACTAGGCC 120
TGTGCCACCCGCTGCCTGCAGGAAGACGCCCGGTCCCGGGCCGGGTTAGCCCCATGGGAA 180
                                          *  p  h  g  n
CGGGGTTCGGTCCGAGCCCGGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCAGCCCACCC
  g  v  r  s  e  p  g  g  r  l  p  e  r  s  l  g  p  a  h  p
CGCGCCGGCGGCCATGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGGAGGAGCT
  a  p  a  a  M  A  G  T  L  D  L  D  K  G  C  T  V  E  E  L
```

Figure 3

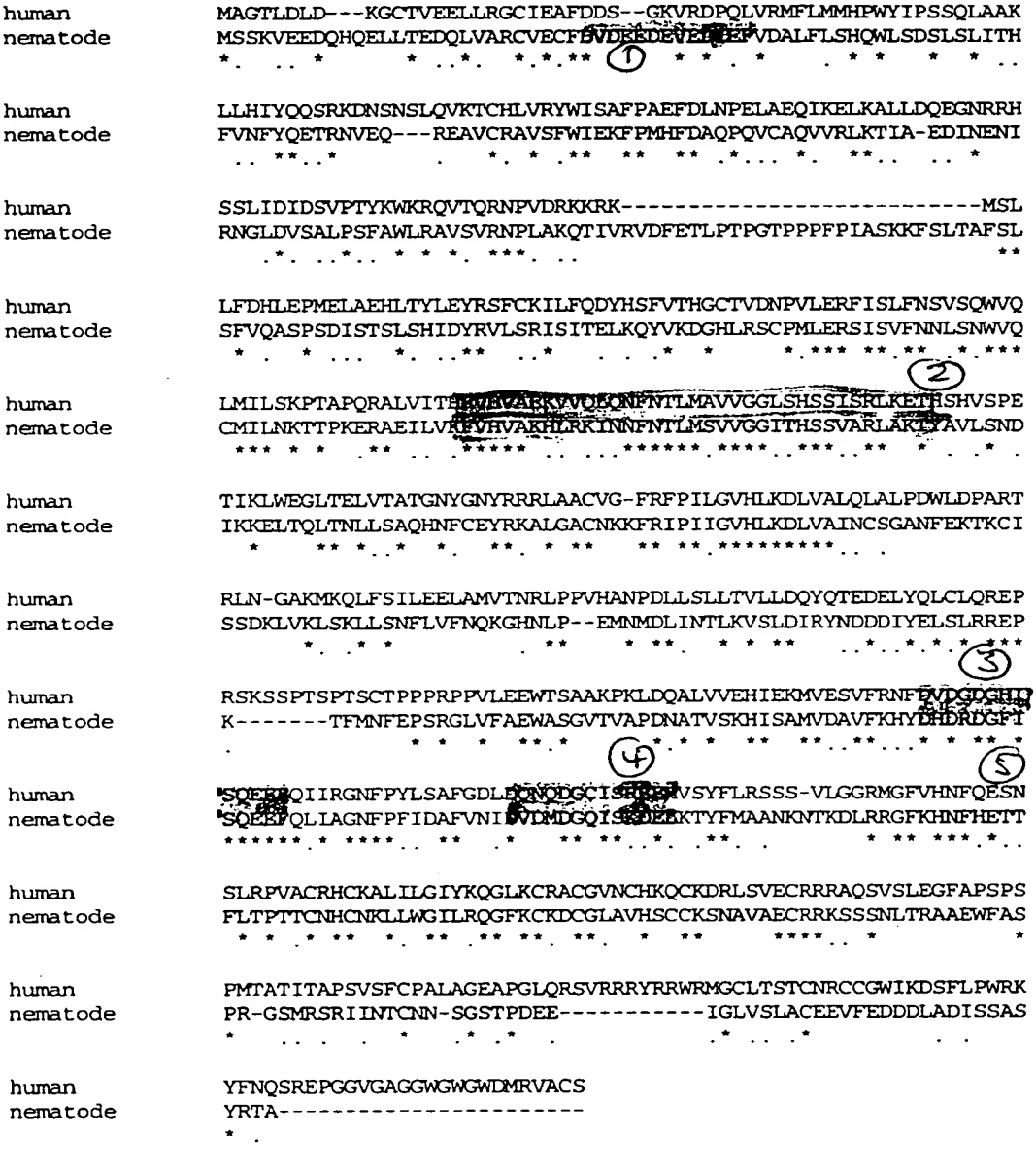

human
nematode
MAGTLDLD---KGCTVEELLRGCIEAFDDS--GKVRDPQLVRMFLMMHPWYIPSSQLAAK
MSSKVEEDQHQELLTEDQLVARCVECFDVDREDEVEEQEFVDALFLSHQWLSDSLSLITH
LLHIYQQSRKDNSNSLQVKTCHLVRYWISAFPAEFDLNPELAEQIKELKALLDQEGNRRH
FVNFYQETRNVEQ---REAVCRAVSFWIEKFPMHFDAQPQVCAQVVRLKTIA-EDINENI
SSLIDIDSVPTYKWKRQVTQRNPVDRKKRK----------MSL
RNGLDVSALPSFAWLRAVSVRNPLAKQTIVRVDFETLPTPGTPPPFPIASKKFSLTAFSL
LFDHLEPMELAEHLTYLEYRSFCKILFQDYHSFVTHGCTVDNPVLERFISLFNSVSQWVQ
SFVQASPSDISTSLSHIDYRVLSRISITELKQYVKDGHLRSCPMLERSISVFNNLSNWVQ
LMILSKPTAPQRALVITFMGVAKKVVQENFNTLMAVVGGLSHSSISRLKETHSHVSPE
CMILNKTTPKERAEILVKFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTYAVLSND
TIKLWEGLTELVTATGNYGNYRRRLAACVG-FRFPILGVHLKDLVALQLALPDWLDPART
IKKELTQLTNLLSAQHNFCEYRKALGACNKKFRIPIIGVHLKDLVAINCSGANFEKTKCI
RLN-GAKMKQLFSILEELAMVTNRLPPVHANPDLLSLLTVLLDQYQTEDELYQLCLQREP
SSDKLVKLSKLLSNFLVFNQKGHNLP--EMNMDLINTLKVSLDIRYNDDDIYELSLRREP
RSKSSPTSPTSCTPPPRPPVLEEWTSAAKPKLDQALVVEHIEKMVESVFRNFEVDGDGHID
K-------TFMNFEPSRGLVFAEWASGVTVAPDNATVSKHISAMVDAVFKHYDHDRDGFI
SQEEIQIIRGNFPYLSAFGDLDGNQDGCISKDEVSYFLRSSS-VLGGRMGFVHNFQESN
SQEEVQLIAGNFPFIDAFVNIDVDMDGQISKDEFKTYFMAANKNTKDLRRGFKHNFHETT
SLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVECRRRAQSVSLEGFAPSPS
FLTPTTCNHCNKLLWGILRQGFKCKDCGLAVHSCCKSNAVAECRRKSSSNLTRAAEWFAS
PMTATITAPSVSFCPALAGEAPGLQRSVRRRYRRWRMGCLTSTCNRCCGWIKDSFLPWRK
PR-GSMRSRIINTCNN-SGSTPDEE----------IGLVSLACEEVFEDDDLADISSAS
YFNQSREPGGVGAGGWGWGWDMRVACS
YRTA----------
1
2
3
4
5

Figure 4

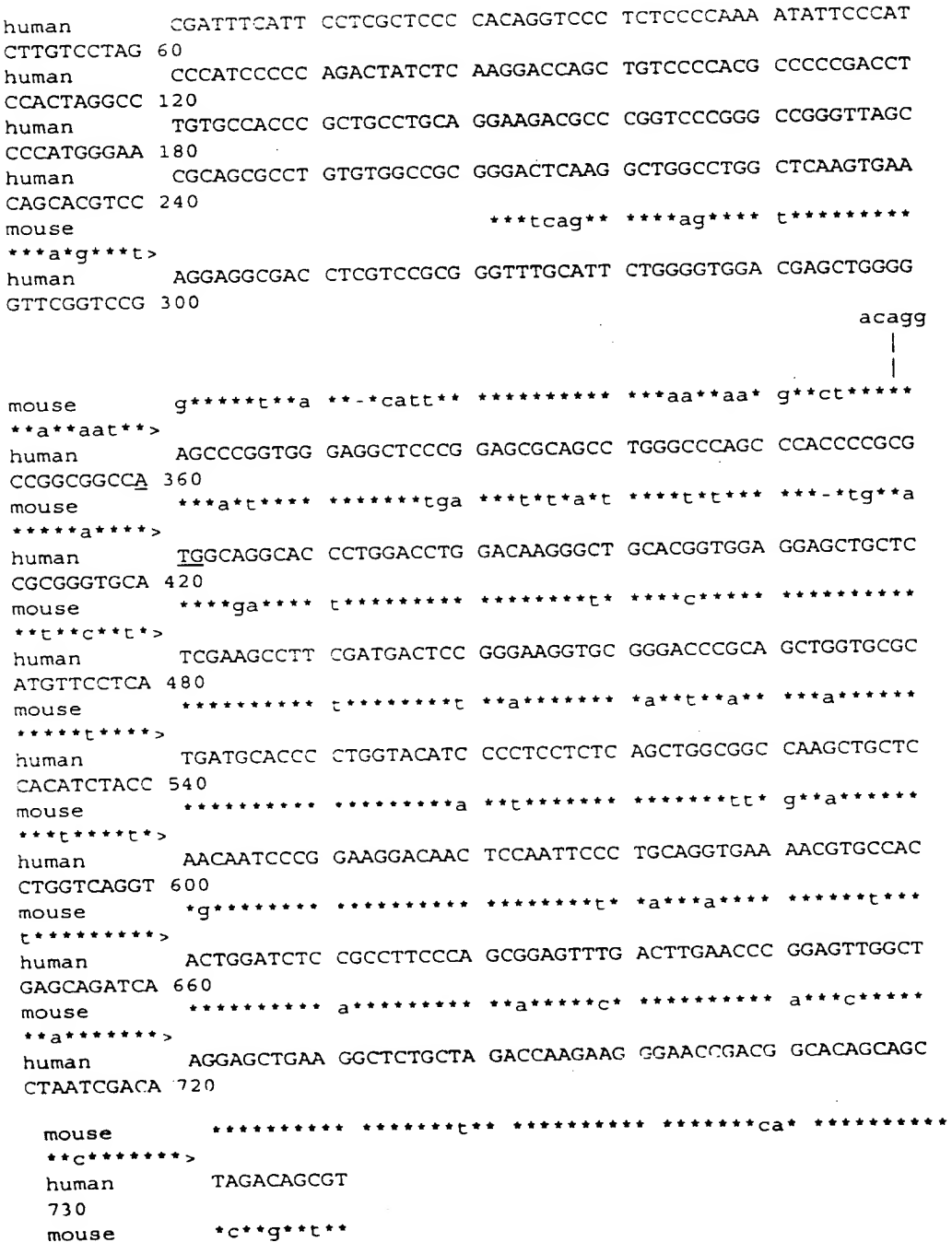

human CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT
CTTGTCCTAG 60
human CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT
CCACTAGGCC 120
human TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAGC
CCCATGGGAA 180
human CGCAGCGCCT GTGTGGCCGC GGGACTCAAG GCTGGCCTGG CTCAAGTGAA
CAGCACGTCC 240
mouse ***tcag** ****ag**** t**********
a*gt>
human AGGAGGCGAC CTCGTCCGCG GGTTTGCATT CTGGGGTGGA CGAGCTGGGG
GTTCGGTCCG 300
acagg
mouse g*****t**a **-*catt** ********** ***aa**aa* g**ct******
aaat**>
human AGCCCGGTGG GAGGCTCCCG GAGCGCAGCC TGGGCCCAGC CCACCCCGCG
CCGGCGGCCA 360
mouse ***a*t***** *******tga ***t*t*a*t ****t*t*** ***-*tg**a
*****a****>
human TGGCAGGCAC CCTGGACCTG GACAAGGGCT GCACGGTGGA GGAGCTGCTC
CGCGGGTGCA 420
mouse ****ga**** t********* ********t* ****c***** **********
tc**t*>
human TCGAAGCCTT CGATGACTCC GGGAAGGTGC GGGACCCGCA GCTGGTGCGC
ATGTTCCTCA 480
mouse ********** t********t **a******* *a**t**a** ***a******
*****t****>
human TGATGCACCC CTGGTACATC CCCTCCTCTC AGCTGGCGGC CAAGCTGCTC
CACATCTACC 540
mouse ********** *********a **t******* ********tt* g**a******
t*t*>
human AACAATCCCG GAAGGACAAC TCCAATTCCC TGCAGGTGAA AACGTGCCAC
CTGGTCAGGT 600
mouse *g******** ********** ********t* *a***a**** ******t***
t**********>
human ACTGGATCTC CGCCTTCCCA GCGGAGTTTG ACTTGAACCC GGAGTTGGCT
GAGCAGATCA 660
mouse ********** a********** **a*****c* ********** a***c*****
a******>
human AGGAGCTGAA GGCTCTGCTA GACCAAGAAG GGAACCGACG GCACAGCAGC
CTAATCGACA 720
mouse ********** *******t** ********** *******ca* **********
c******>
human TAGACAGCGT
730
mouse *c**g**t**

Figure 5

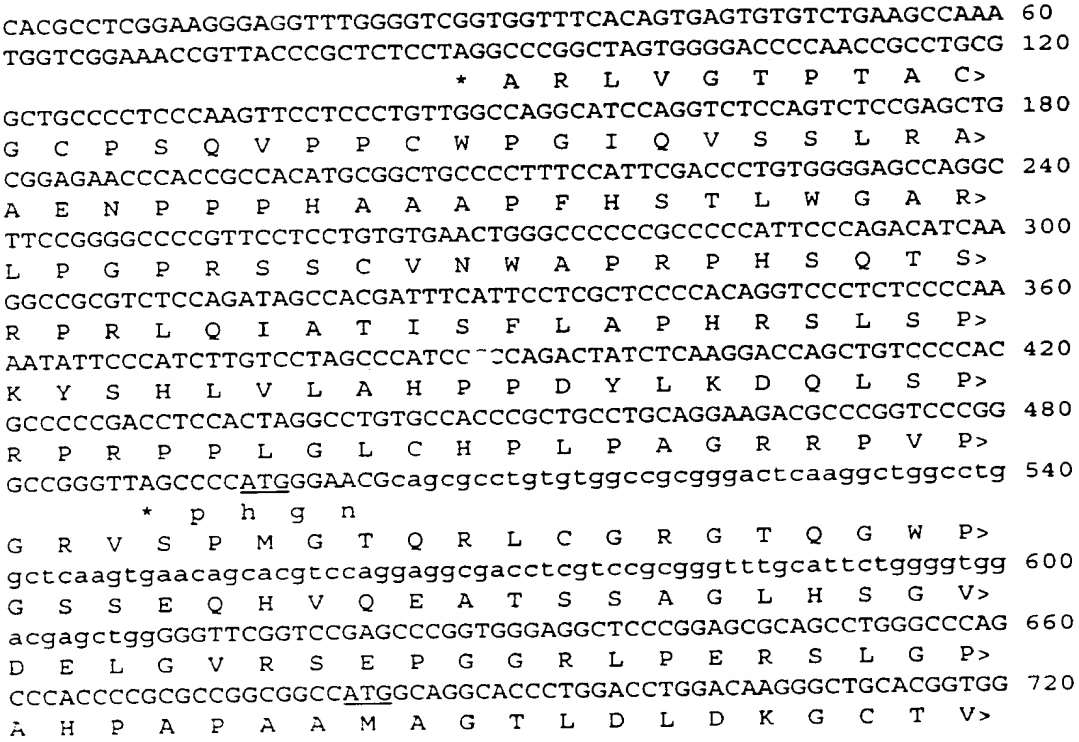
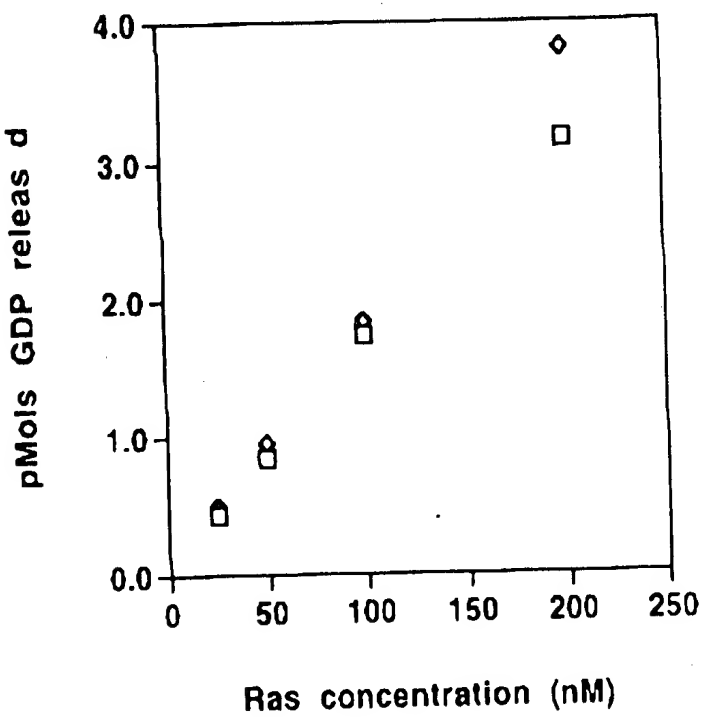

CACGCCTCGGAAGGGAGGTTTGGGGTCGGTGGTTTCACAGTGAGTGTGTCTGAAGCCAAA 60
TGGTCGGAAACCGTTACCCGCTCTCCTAGGCCCGGCTAGTGGGGACCCCAACCGCCTGCG 120
* A R L V G T P T A C>
GCTGCCCCTCCCAAGTTCCTCCCTGTTGGCCAGGCATCCAGGTCTCCAGTCTCCGAGCTG 180
G C P S Q V P P C W P G I Q V S S L R A>
CGGAGAACCCACCGCCACATGCGGCTGCCCCTTTCCATTCGACCCTGTGGGGAGCCAGGC 240
A E N P P P H A A A P F H S T L W G A R>
TTCCGGGGCCCCGTTCCTCCTGTGTGAACTGGGCCCCCCGCCCCCATTCCCAGACATCAA 300
L P G P R S S C V N W A P R P H S Q T S>
GGCCGCGTCTCCAGATAGCCACGATTTCATTCCTCGCTCCCCACAGGTCCCTCTCCCCAA 360
R P R L Q I A T I S F L A P H R S L S P>
AATATTCCCATCTTGTCCTAGCCCATCC CCAGACTATCTCAAGGACCAGCTGTCCCCAC 420
K Y S H L V L A H P P D Y L K D Q L S P>
GCCCCCGACCTCCACTAGGCCTGTGCCACCCGCTGCCTGCAGGAAGACGCCCGGTCCCGG 480
R P R P P L G L C H P L P A G R R P V P>
GCCGGGTTAGCCCCATGGGAACGcagcgcctgtgtggccgcgggactcaaggctggcctg 540
* p h g n
G R V S P M G T Q R L C G R G T Q G W P>
gctcaagtgaacagcacgtccaggaggcgacctcgtccgcgggtttgcattctggggtgg 600
G S S E Q H V Q E A T S S A G L H S G V>
acgagctggGGGTTCGGTCCGAGCCCGGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCAG 660
D E L G V R S E P G G R L P E R S L G P>
CCCACCCCGCGCCGGCGGCCATGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGG 720
A H P A P A A M A G T L D L D K G C T V>

Figure 6


4.0
3.0
2.0
1.0
0.0
pMols GDP releas d
0
50
100
150
200
250
Ras concentration (nM)